

REGRESIE LINIARĂ TIP „SPLINE” UTILIZATĂ ÎN EVALUAREA ACTIVELOR FINANCIARE¹

Doctorand Liviu GEAMBAȘU
Doctorand Ionel JIANU
lector univ. dr. Cristina GEAMBAȘU
Academia de Studii Economice București

SPLINE LINEAR REGRESSION USED FOR EVALUATING FINANCIAL ASSETS¹

PhD student Liviu GEAMBAȘU
PhD student Ionel JIANU
Lecturer PhD. Cristina GEAMBAȘU
Academy of Economic Studies Bucharest

Rezumat:

Principală preocupare a actorilor piețelor financiare a fost și rămâne determinarea cât mai exactă a evoluției cursurilor activelor financiare. În acest sens au fost scrise multe articole științifice și au fost dezvoltate multe modele matematice și statistice care să ajute la identificarea trendurilor de evoluție a activelor financiare.

Dacă până de curând modelele liniare simple erau larg răspândite datorită simplității în utilizare, criza financiară, care a afectat economia mondială începând din 2008, a pus în evidență necesitatea adaptării modelelor matematice la variațiile economiei. Un model simplu de utilizat și totuși adaptat realității economice este reprezentat de regresii liniare tip „spline”. Acest tip de regresie liniară menține continuitatea funcției de regresie, dar împarte datele studiate în intervale de valori cu caracteristici cât mai omogene. Se evidențiază caracteristicile fiecărui interval și evoluția pieței de-a lungul tuturor intervalelor, cu erori standard de evaluare mai mici.

Primul obiectiv al prezentului articol este prezentarea elementelor teoretice ale regresiei tip „spline”, cu referințe la principalele lucrări științifice apărute pe plan internațional și național. Cel de al doilea obiectiv este reprezentat de aplicarea modelului pe datele oferite de Bursa de Valori București.

Cuvinte cheie: rentabilitate valori mobiliare, regresie spline, model evaluare active, subregiuni, abatere standard a erorilor.

1. Introducere

Evaluarea corectă a evoluției piețelor de capital și a prețurilor activelor financiare este o preocupare esențială a finanțistilor teoreticieni dar și practicieni din lumea întreagă. Criza financiară care afectează economia mondială începând din anul 2008 a readus problema evaluării activelor printre

Abstract:

One of the most important preoccupations of financial markets participants was and still is the problem of determining more precise the trend of financial assets prices. For solving this problem there were written many scientific papers and were developed many mathematical and statistical models in order to better determine the financial assets price trend.

If until recently the simple linear models were largely used due to their facile utilization, the financial crises that affected the world economy starting with 2008 highlight the necessity of adapting the mathematical models to variation of economy. A simple to use model but adapted to economic life realities is the spline linear regression. This type of regression keeps the continuity of regression function, but split the studied data in intervals with homogenous characteristics. The characteristics of each interval are highlighted and also the evolution of market over all the intervals, resulting reduced standard errors.

The first objective of the article is the theoretical presentation of the spline linear regression, also referring to scientific national and international papers related to this subject. The second objective is applying the theoretical model to data from the Bucharest Stock Exchange.

Key words: assets return, spline regression, asset evaluation model, subsets of data, standard deviation of errors.

1. Introduction

Correct evaluation of capital market and financial assets prices evolution is an essential preoccupation for both theoretical and business finance people all around the world. The financial crisis that affects the world economy starting from 2008 had put the subject of assets evaluation through the

principalele preocupări ale economiștilor.

Regresiile spline sunt modele relativ ușor de creat și aplicat, oferind rezultate foarte bune în estimarea datelor. Caracteristica acestor regresii este reprezentat de divizarea populației statistice în grupe omogene, cercetate prin intermediul unei singure funcții generale, dar cu caracteristici proprii fiecărei regiuni. Un alt avantaj al regresiiilor spline este capacitatea de a prezenta evoluția anterioară a prețurilor, ajutând la înțelegerea mai bună a modului de funcționare a pieței și de manifestare a investitorilor.

Articolul are în continuare următoarea dezvoltare: capitolul 2 prezintă utilizarea regresiei spline în literatura de specialitate internațională și câteva domenii de utilizare, capitolul 3 care este o descriere teoretică a conceptului de regresie spline, incluzând și aspecte forte și sensibile ale regresiiilor spline și modalități de stabilire a intervalelor. Capitolul 4 prezintă datele utilizate, modelul testat și rezultatele obținute prin aplicarea modelului. Studiul se termină cu capitolul 5 care prezintă concluziile și direcțiile viitoare de cercetare.

2. Regresie spline în literatura de specialitate internațională

Problematika studiată preocupă domenii de cercetare multiple: matematica – funcții de predicție multivariate, statistica – regresie multiplă non-parametrică, ingineria și dezvoltarea software – rețele neuronale.

Friedman (1991) dezvoltă modelul „recursive partitioning regression”, realizând o funcție de predicție cumulativă care reduce erorile standard de estimare pe baza datelor observate empiric în timp pentru variabila dependentă și oferă o înțelegere a modului de evoluție a variabilei dependente. Funcția cumulativă reunește funcțiile adaptate fiecărei subregiuni din spațiul cercetat, iar partiționarea în subregiuni se realizează prin împărțirea spațiului inițial în două subregiuni, și recursiv a fiecărei subregiuni în alte două subregiuni, proces repetat până la crearea unui număr de subregiuni optim. Funcția

main concerns of the economists.

The spline regressions are models easy to build and applied, offering good results in data approximation. The characteristic of the regression is the split of statistical population in homogeneous groups, researched through a single general function, that have characteristics proper to each sub-region. Another advantage of spline regression is its capacity of reveal historical evolution of prices, helping in a better understanding of the way the markets function and of investor manifestation.

The paper continues with the following: chapter 2 presents usage of spline regression in the international specialty literature and some of the domain where it was used, chapter 3 is a theoretical description of the spline regression concept, including also some pros and cons of spline regression and modalities of establishing intervals. Chapter 4 presents utilized data, tested model and the results obtained after applying the model. The paper ends with chapter 5 that presents conclusion and future directions of research.

2. Spline regression in the international specialty literature

The issue studied is researched in several scientific domains such as: mathematics – multivariate forecast functions, statistics – non-parametric multiple regression, engineering and software development – neuronal networks.

Friedman (1991) expands the „recursive partitioning regression” model and obtains a prediction cumulative function that reduces the forecast standard errors based on empirical observed data for dependent variable and offers a better understanding of the evolution of dependent variable. The cumulative function reunites function adapted to each sub-region from the researched space, partitioning in sub-regions is realized through splitting the initial space in two sub-regions, process repeated until the total number of obtained sub-regions is optimal. The function proper to each sub-

proprie subregiunii poate scoate în evidență influența regională mai puternică a anumitor variabile independente asupra variabilei dependente. Sunt incluse astfel într-o singură funcție toate funcțiile aferente subregiunilor și apoi anulate funcțiile subregionale în dependență cu valorile variabilei independente. Se poate asigura astfel caracterul de continuitate a funcției generale, inclusiv la limitele subregiunilor.

Funcțiile non-parametrice sunt folosite și în alte domenii pentru determinarea legilor proceselor economice (modelele non-parametrice se adaptează evoluției în timp a datelor și nu stabilesc forma dependenței apriori). Richard Blindell, Xiaohong Chen și Dennis Kristensen (2007) au studiat legătura dintre cererea de bunuri și evoluția bugetelor gospodăriilor. Modelul creat de ei, adaptat propriei cercetări, urmărește aceeași linie directoare – minimizarea erorilor de previziune ale modelului relativ la datele empirice prin utilizarea unei funcții continue pe subseturi de date rezultată din însumarea funcțiilor aferente subseturilor de date într-o funcție generală și evidențierea caracteristicilor proprii subseturilor.

Priya Ranjan și Justin L. Tobias (2007) folosesc modelul tobit (propus de James Tobin în 1958) pentru estimarea legăturilor comerciale dintre două state luând în calcul PIB-ul celor două state și vecinătatea acestora. Modelul tobit este un model asemănător, presupunând valoarea nulă a unei funcții dacă nu este îndeplinită condiția ca variabila independentă să fie inclusă într-o arie de valori.

Robert F. Engle și Jose Gonzalo Rangel (2008) au studiat volatilitatea pieței financiare în corelație cu evoluția macroeconomică a 50 de state, pe baza unei regresii spline-GARCH în descrierea trendului cu frecvență redusă de modificare a volatilității variabilei macroeconomice în timp. Cei doi autori au creat o funcție continuă, pentru descrierea influenței variabilelor independente asupra celei dependente și pentru reliefarea evoluției variabilei dependente în timp (evoluție

region emphasizes the regional influence of some independent variables on dependent variable. There are included in a single function all the functions of the sub-regions and nullified the sub-regions functions depending on independent variables values. There can be ensured the continuity character of the general function, and also for the sub-regions limits.

Non-parametric functions are used in many other domains for determining the economic processes laws (non-parametric models adapts the evolution in time of data and do not have a dependency form established from beginning). Richard Blindell, Xiaohong Chen and Dennis Kristensen (2007) used a non-parametric model in order to study the relation between good demands and household budgets. The model created by them, adapted to their own research, follows the same research line – minimizing the forecast error of the model related to the empirical data using a continuous function on subsets of data resulted from adding the subsets functions in a single general function and highlighting the characteristics of each subset.

Priya Ranjan and Justin L. Tobias (2007) use the tobit model (developed by James Tobin in 1958) for estimation of commercial trades between two states counting the GDP of the two states and their neighborhood. The tobit model is similar, implying nullifying a function if there is not accomplished the condition that an independent variable to be in a set of values.

Robert F. Engle and Jose Gonzalo Rangel (2008) studied the volatility of the financial market in correlation with macroeconomic evolution for 50 states, using a GARCH-spline regression model for describing the trend with reduce frequency of the macroeconomic variable volatility in time. The authors created a function continuous appropriate for description of independent variables influence over the dependent variable and to highlight the evolution of dependent variable in time (evolution of GDP, yield rate, inflation,

semnificativă pentru PIB, rata de dobândă, inflația, numărul de companii listate la bursă, mărimea pieței financiare relativ la PIB).

Utilizarea regresiiilor spline este larg răspândită și în alte domenii de specialitate, fiind folosite în analize privind sănătatea publică, cercetări demografice, etc.; modelele create fiind adaptate problemelor cercetate, urmărind să aproximeze cât mai exact realitatea empirică, pe baza unor regresii ușor de creat, folosit și adaptate evoluției în timp a datelor.

3. Descriere teoretică a regresiei spline

3.1. Forma teoretică a unei regresii spline

Criza economică mondială impune ajustarea modelelor create anterior la noii parametri de evoluție a pieței; criza financiară a modificat sensurile de evoluție a prețurilor activelor și modul comportare al investitorilor în tranzacționarea activelor, astfel că orice regresie care utilizează date istorice nu poate fi folosită în evaluarea tendințelor actuale. Au fost reevaluate tendințele dar și înclinațiile spre consum, economisire, investire și parametrii regresiiilor anterioare – parametrii stabiliți pe baza datelor istorice.

Regresia spline este foarte utilă în descrierea evoluțiilor cu trenduri contrare pe intervale de timp succesive, permițând crearea unei ecuații generale de evoluție, cu unul sau mai mulți factori de influență, cu forme de determinare mai simple sau mai complexe. Caracteristica principală a acestui tip de regresie constă în divizarea populației studiate în grupe cu caracteristici omogene, menținând însă forma unitară a funcției de regresie.

Regresia spline utilizează intervale de evoluție omogene pentru variabilele independente și determină pante de regresie specifice fiecărui interval, generând valori calculate ale variabilei dependente mult mai apropiate de valorile reale observate empiric.

Regresia spline are următoarea formă:

number of companies listed on the stock exchange, volume of financial market compared with the GDP).

Usage of spline regression is largely spread in other domains of activity, being also used in analyses regarding public health, demographic research, so on; models developed being adapted to the researched themes, intended to better approximate the empirical reality based on regressions easy to create, use and adaptive to evolution of data in time.

3. Theoretical presentation of spline regression

3.1. Theoretical form of the spline regression

The world economic crises impose the adjusting of the models created in the past to the new markets evolution parameters; the financial crises modified the evolution sense for assets prices and the investors' behavior in trading assets, so that any regression using historical data cannot be used in evaluation of present time trends. There were evaluated trends and marginal rates for consume, economy, investments as well former regressions parameters – parameters established based on historical data.

The spline regression is used in description of trends with opposite evolution on successive time intervals, permitting the creation of a generalized equation, having one or more influencing factors and simple or complex deterministic forms. The main characteristic of this type of regression is the split of researched population in homogeneous groups, but keeping the unitary form of the regression function.

The spline regression uses homogenous intervals for independent variables and determines the regression slope for each interval, generating computed values for the dependent variable closer to the real empirical observations.

The spline regression has the

$$\hat{y} = \sum_{i=1}^n a_i F_i(x) \quad (1)$$

cu a_i este o constantă diferită specifică fiecărui interval iar $F_i(x)$ este o funcție de forma:

$$F_i = \max(b_i, c_i) = \begin{cases} b_i, & \text{daca } b_i > c_i \\ c_i, & \text{daca } b_i < c_i \end{cases} \quad (2)$$

cu b_i este o constantă iar c_i o funcție care reflectă relația dintre variabila independentă și cea dependentă – exemplu: relație liniară, o progresie, un logaritm, etc., având una sau mai multe variabile independente. $F_i = 1$ când a_i reprezintă termenul liber (engl. intercept)

Pentru o regresie liniară cu o singură variabilă independentă, funcția $F_i(x)$ poate avea una din următoarele forme: $\max(0, x_i - d_i)$ sau $\max(0, d_i - x_i)$, unde x_i este variabila independentă iar d_i este constanta specifică fiecărui interval. În acest caz:

$$\hat{y} = \sum_{i=1}^n a_i F_i(x) = \sum_{i=1}^n a_i \max(0, x_i - d_i) \quad (3)$$

Intervalele de regresie vor fi determinate de valorile pentru care $F_i(x) = \max(0, x_i - d_i)$ devine nulă ($F_i(x) = 0$), reprezentând una dintre provocările acestui tip de funcție. Împărțirea în intervale este limitată doar de numărul de intervale dorit de utilizator și de relevanța fiecărui interval – împărțirea în prea multe intervale face ca informația rezultată în fiecare interval să devină nerelevant.

Cercetătorul poate crea o funcție $F_i(x)$ continuă pe întreg intervalul cercetat, funcție compusă din mai multe segmente, fiecare cu propria pantă de regresie.

3.2. Elemente pro și contra

Deși regresia spline reprezintă o abordare utilă în descrierea evoluției evenimentelor economice, există elemente pro și contra, cu caracter general; în funcție

$$\text{following form: } \hat{y} = \sum_{i=1}^n a_i F_i(x) \quad (1)$$

having a_i a different constant specific to each interval and $F_i(x)$ a function like:

$$F_i = \max(b_i, c_i) = \begin{cases} b_i, & \text{daca } b_i > c_i \\ c_i, & \text{daca } b_i < c_i \end{cases} \quad (2)$$

having b_i a constant and c_i a function that reflects the relation between the independent variable and the dependent variable – example: a linear relation, an exponential function, a logarithm function, and so on; having one or more independent variables. $F_i = 1$ if a_i represents the intercept.

For a linear regression with a single independent variable, the $F_i(x)$ function may have one of the following forms: $\max(0, x_i - d_i)$ or $\max(0, d_i - x_i)$, having x_i as independent variable and d_i as a constant specific to each interval. In this case the regression has the form:

$$\hat{y} = \sum_{i=1}^n a_i F_i(x) = \sum_{i=1}^n a_i \max(0, x_i - d_i) \quad (3)$$

The regression intervals will be determined by the values for which $F_i(x) = \max(0, x_i - d_i)$ is null ($F_i(x) = 0$), representing one of the tricky problems of this regression. Splitting in intervals is limited only by the number of intervals the user wants and by the relevance of each interval – splitting in too many intervals determine each interval to become irrelevant.

The researcher can create a function $F_i(x)$ continuous on the entire studied interval, formed of multiple intervals, each with its own regression slope.

3.2. Pros and cons issues

Although the spline regression is a useful approach in describing the economic events evolutions, there are pros and cons

de tema cercetată și de rezultatele urmărite, existând și alte argumente pro / contra sau unele sunt aplicabile.

Elementul principal care susține utilizarea regresiei spline este flexibilitatea acesteia în descrierea evoluției fenomenelor pe intervale cu evoluții diferite. Se obține o mai bună determinare a valorilor variabilei dependente, și se observa modul în care variabilele independente au evoluat în timp și modificarea pantei de regresie între intervale.

Pentru o înregistrare empirică vom avea o valoare a variabilei determinate (notată \hat{y}_j^i - \hat{y}_j aparținând intervalului i , x_j^i fiind variabila independentă):

$$\hat{y}_j^i = a_1 + a_i \max(0, x_j^i - d_i) \quad (4)$$

Pentru un anumit interval i determinat, ecuațiile de forma $\max(0, x_k - d_k)$ unde $k \neq i$ sunt egale cu zero – fiind ecuații aferente restului intervalelor.

Valorile erorilor standard ale regresiei sunt mai mici în cazul regresiei spline decât în cazul unei regresii liniare.

3.3. Determinarea numărului de intervale

Unul din aspectele importante ale regresiei spline este determinarea numărului de intervale adecvat pentru studierea populației observate. În literatura de specialitate există mai multe moduri de abordare: stabilirea unui număr fix de intervale, stabilirea numărului de intervale pe baza trendurilor de evoluție (perioade de creștere și descreștere), stabilirea intervalelor pe baza omogenității populației care compune fiecare interval, etc.

În situația în care pentru determinarea intervalelor nu există un criteriu clar, sau în situația în care numărul intervalelor ar crește considerabil, îngreunând atât procesul de obținere a rezultatelor sau când interpretarea datelor devine imposibilă datorită numărului mare de intervale există diverse criterii de selecție a numărului optim de intervale.

Un criteriu de selectare a numărului

aspecte, with general character; considering the researched theme and the expected results there can be others pro / con arguments or some are not applicable.

The main issue that sustains the usage of spline regression is its flexibility in describing phenomenon evolution on intervals with various trends. There can be obtained a better determination of values for the dependent variables, and also can be observed the manner in which the independent variables evaluate over time and modification of slopes between intervals.

For an empirical observed value there will be the value of dependent variable (noted \hat{y}_j^i - \hat{y}_j for interval i , x_j^i being the independent variable):

$$\hat{y}_j^i = a_1 + a_i \max(0, x_j^i - d_i) \quad (4)$$

For a specific interval i , the equations $\max(0, x_k - d_k)$ with $k \neq i$ are equal with zero – there are equations proper to other intervals.

The values of standard errors of the regressions are smaller for the spline regression then for a linear regression.

3.3. Determining the number of intervals

One of the important issue of spline regression is how to determine the adequate number of intervals for studying the observed population. In the specialty literature exists couple of methods of approaching this issue: establish a fix number of intervals, establish a number of intervals based on evolution trends (intervals with growth and intervals with decrease for trends), establish intervals based on the homogeneity of the population composing each interval, so on.

If determining intervals does not follow a precise criterion or if the number of intervals would grow exponentially, making difficult the process of obtaining valid results or the interpretation of data is impossible due to increase number of

optim de intervale este GCV (engleză Generalised Cross Validation) care penalizează abaterea erorilor standard cu numărul de intervale create. Elementul central al GCV este reprezentat de minimizarea abaterii erorilor standard ale valorilor obținute din calcule relativ la valorile observate empiric; fiind motivul care stă la baza tuturor cercetărilor – obținerea de valori teoretice cât mai apropiate de cele observate empiric. Abaterea erorilor scade pe măsură ce numărul intervalelor crește, însă creșterea numărului intervalelor induce probleme de determinare a rezultatelor și de interpretarea acestora într-un timp util și rezonabil.

GCV penalizează criteriul de minimizare al abaterilor erorilor prin includerea numărului de intervale creat, pentru a crea un echilibru între divizarea excesivă a populației cercetate și optimizarea soluției cercetate.

$$GCV = \frac{1}{N} \times \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\left[1 - \frac{M + d * M + 1}{N}\right]^2} \quad (5)$$

unde M reprezintă numărul de parametri ai funcției de regresie iar d este un coeficient care asigură planeitatea funcției. Valorile usuale pentru d sunt 0 pentru unele studii și 2 pentru studiile care urmăresc și criteriul de planeitate al GCV (uneori chiar 3 – asigurând un nivel mai mare de planeitate).

Utilizarea unui număr corect de intervale este considerată ca aspect care asigură funcției de regresie spline flexibilitate și în egală măsură putere de reprezentare a informațiilor privind populația cercetată.

4. Modelul de regresie cercetat

4.1. Datele utilizate

Folosim cursul zilnic la închidere al acțiunilor listate la Bursa de Valori București, înregistrat în perioada 15 mai 2009 – 11 noiembrie 2009, respectiv 128 de cotații zilnice.

intervals, there are some statistical criterion for selecting the optimal number of intervals.

A criterion of selecting the optimal number of intervals is GCV (Generalized Cross Validation) that penalized the deviation of errors with the increase in number of intervals. The central figure of GCV is represented by minimization of variance of errors from computed values comparing with empirical observed values; this is the reason for all research – obtaining theoretical values closer to empirical observed ones. Deviation of errors decrease if number of intervals increase, but increase in number of intervals induce problems of determining results and in interpreting the results in a reasonable time and with reasonable costs.

GCV penalizes the minimization criterion of error variation through including the increase in number of intervals, in order to ensure an equilibrium between exponential splitting of population researched and optimization of obtained solution.

$$GCV = \frac{1}{N} \times \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\left[1 - \frac{M + d * M + 1}{N}\right]^2} \quad (5)$$

having M as the number of parameters of the regression function, and d is a coefficient that insure the smoothness of the function. Usual values for d are 0 for some studies and 2 for papers that take into account also the smoothness criterion for GCV (sometimes 3 ensuring a better smoothness).

Using a right number of intervals ensure is consider to ensure the spline regression flexibility and give the strength of representing data related to researched population.

4. Researched regression model

4.1. Utilized data

We use daily closing price of shares

Piața de capital din România are o lichiditate redusă, de aceea utilizăm acele acțiuni care au un număr de zile de cotaie cât mai apropiat de numărul de zile maxim. Astfel, vom folosi cotațiile următoarelor titluri: SIF5, SIF2, TLV, BRD, SIF3, SNP, SIF4, SIF1, TGN.

Pentru reprezentarea evoluției pieței vom folosi valorile indicelui bursier BET-XT –care reprezintă evoluția pieței bursiere în ansamblu prin includerea tuturor acțiunilor din piață.

Împărțim întregul interval studiat în 6 subregiuni arbitrar definite, de la ziua 15 a lunii X până la ziua 14 a lunii X+1.

4.2. Modelul regresiei spline studiat

Vom testa utilizarea regresiei spline pentru a determina influența evoluției acțiunilor asupra rentabilității pieței.

Regresia liniară are următoarea formă:

$$\hat{R}_{Mz} = a_0 + \sum_{t=1}^9 b_t \times R_{tz} \quad (6)$$

unde t reprezintă titlurile incluse în studiu, z reprezintă ziua evaluării.

Regresia spline utilizată va avea următoarea formă:

$$\hat{R}_{Mzj} = \sum_{j=1}^6 \left[\frac{\left(\alpha_{0j} + \sum_{t=1}^9 b_{tj} \times R_{tzj} \right) \times \left(1 - \frac{1}{\left(Z_{Mj} - z_j \right) \times \left(Z_{mj} - z_j \right)} \right)^x}{\max \left[0, \left(Z_{Mj} - z_j \right) \times \left(Z_{mj} - z_j \right) \right]} \right] \quad (7)$$

unde \hat{R}_{Mzj} reprezintă rentabilitatea pieței calculată în ziua z aparținând subregiunii j (populația cercetată este împărțită în 6 subregiuni), a_{0j} sunt constante proprii subregiunilor, b_{tj} reprezintă influența titlului t pentru subregiunea j , iar R_{tzj} reprezintă

listed on Bucharest Stock Exchange, between May 15, 2009 and November 11, 2009 – a total of 128 quotation days.

The capital market in Romania has a reduce liquidity, so we use only those shares that have a number of quotation days closer to the maxim of 128 possible quotation days; the titles are: SIF5, SIF2, TLV, BRD, SIF3, SNP, SIF4, SIF1, TGN.

For evolution of market we shall use the values of stock exchange index BET-XT – that represents the evolution of stock market including all shares listed.

We split the entire interval in 6 sub-regions arbitrary defined, from day 15 of month X to day 14 of month X+1.

4.2. Researched spline regression model

We shall test the spline regression for determining the influence of shares evolution over market return.

The linear regression has the following form:

$$\hat{R}_{Mz} = a_0 + \sum_{t=1}^9 b_t \times R_{tz} \quad (6)$$

where t represents the titles included in the study and z is the day for computed data.

The spline regression will have the following form:

$$\hat{R}_{Mzj} = \sum_{j=1}^6 \left[\frac{\left(\alpha_{0j} + \sum_{t=1}^9 b_{tj} \times R_{tzj} \right) \times \left(1 - \frac{1}{\left(Z_{Mj} - z_j \right) \times \left(Z_{mj} - z_j \right)} \right)^x}{\max \left[0, \left(Z_{Mj} - z_j \right) \times \left(Z_{mj} - z_j \right) \right]} \right] \quad (7)$$

where \hat{R}_{Mzj} represents market return computed for day z in the sub-region j (the population is split in 6 sub-regions), a_{0j} are constants proper to each sub-region, b_{tj} represents the influence of title t for sub-

rentabilitatea titlului t în ziua z din cadrul subregiunii j , Z_{Mj} - ultima zi din intervalul j , Z_{mj} - prima zi din intervalul j , z_j - ziua curentă din intervalul j .

Eroarea va fi determinată prin compararea valorii empirice a lui R_{ij} cu valoarea calculată a rentabilității zilnice \hat{R}_{ij} :

$\varepsilon_{ij} = R_{ij} - \hat{R}_{ij}$. Pentru ambele regresii se va calcula abaterea standard a erorilor: $\sigma^2(\varepsilon_i) = \frac{(\varepsilon_i - \bar{\varepsilon}_i)^2}{T-1}$ (8)

4.3. Rezultatele cercetării

Pentru regresia liniară obținem următoarele date:

Tabelul 1. Rezultatele regresiei liniare

	BET-C	BRD	SIF1	SIF2	SIF3	SIF4	SIF5	SNP	TGN	TLV
Average	0.31%	0.58%	0.31%	0.40%	0.43%	0.18%	0.31%	0.24%	0.01%	0.62%
b_t		1.25	1.33	1.53	1.38	1.20	1.48	1.30	0.65	1.04

Coeficientul $a_0 = -3.68\%$, iar $\sigma^2(\varepsilon_i) = 0.0853$

Pentru regresia spline înregistrăm următoarele date, cu $\sigma^2(\varepsilon_i) = 0.07783$:

Tabelul 2. Coeficientii de regresie spline ai acțiunilor pe fiecare subregiune (SR)

b_{ij}	BRD	SIF1	SIF2	SIF3	SIF4	SIF5	SNP	TGN	TLV
SR1	1.34	1.63	1.66	1.54	1.43	1.64	1.34	0.79	0.96
SR2	0.86	1.82	1.59	1.73	1.63	1.71	1.55	0.59	1.15
SR3	1.16	1.10	1.47	1.12	0.97	1.28	1.32	0.51	1.87
SR4	1.44	1.04	1.41	1.28	1.03	1.28	1.20	0.75	1.01
SR5	1.25	1.23	1.59	1.22	1.10	1.58	1.38	0.63	0.75
SR6	1.35	1.26	1.60	1.38	1.30	1.52	1.08	0.64	0.63

Tabelul 3. Rentabilitățile acțiunilor pe fiecare subregiune (SR)

Average	BRD	SIF1	SIF2	SIF3	SIF4	SIF5	SNP	TGN	TLV
SR1	1.27%	1.17%	1.04%	1.13%	0.67%	0.87%	1.42%	0.36%	1.17%
SR2	-0.29%	-0.72%	-0.58%	-0.63%	-0.31%	-0.78%	-0.67%	-0.36%	-0.61%
SR3	1.67%	1.24%	1.61%	1.65%	0.76%	1.66%	0.74%	0.38%	1.26%
SR4	0.29%	0.52%	0.59%	0.99%	0.37%	0.29%	-0.22%	-0.18%	0.62%
SR5	0.21%	0.17%	0.23%	0.07%	0.22%	0.09%	0.36%	-0.04%	0.41%
SR6	0.37%	-0.50%	-0.54%	-0.67%	-0.66%	-0.29%	-0.10%	-0.10%	0.98%

Tabelul 4. Coeficientul a_0 , rentabilitatea pieței și variația erorilor pe fiecare subregiune (SR)

region j , R_{tj} represents the return of title t from day z in sub-region j , Z_{Mj} - last day from sub-region j , Z_{mj} - first day from sub-region j , z_j - current day in sub-region j .

The error will be determined by comparing empirical value of R_{ij} with the computed value of daily return \hat{R}_{ij} :

$\varepsilon_{ij} = R_{ij} - \hat{R}_{ij}$. For both regressions there will be calculated the variance of error: $\sigma^2(\varepsilon_i) = \frac{(\varepsilon_i - \bar{\varepsilon}_i)^2}{T-1}$ (8)

4.3. Research results

For linear regression we obtained the following data:

Table 1. Linear regression results

	BET-C	BRD	SIF1	SIF2	SIF3	SIF4	SIF5	SNP	TGN	TLV
Average	0.31%	0.58%	0.31%	0.40%	0.43%	0.18%	0.31%	0.24%	0.01%	0.62%
b_t		1.25	1.33	1.53	1.38	1.20	1.48	1.30	0.65	1.04

coefficient $a_0 = -3.68\%$, and $\sigma^2(\varepsilon_i) = 0.0853$

For the spline regression we the following data, with $\sigma^2(\varepsilon_i) = 0.07783$:

Table 2. Spline regression coefficients for each share and for each sub-region

b_{ij}	BRD	SIF1	SIF2	SIF3	SIF4	SIF5	SNP	TGN	TLV
SR1	1.34	1.63	1.66	1.54	1.43	1.64	1.34	0.79	0.96
SR2	0.86	1.82	1.59	1.73	1.63	1.71	1.55	0.59	1.15
SR3	1.16	1.10	1.47	1.12	0.97	1.28	1.32	0.51	1.87
SR4	1.44	1.04	1.41	1.28	1.03	1.28	1.20	0.75	1.01
SR5	1.25	1.23	1.59	1.22	1.10	1.58	1.38	0.63	0.75
SR6	1.35	1.26	1.60	1.38	1.30	1.52	1.08	0.64	0.63

Table 3. Shares return for each sub-region (SR)

Average	BRD	SIF1	SIF2	SIF3	SIF4	SIF5	SNP	TGN	TLV
SR1	1.27%	1.17%	1.04%	1.13%	0.67%	0.87%	1.42%	0.36%	1.17%
SR2	-0.29%	-0.72%	-0.58%	-0.63%	-0.31%	-0.78%	-0.67%	-0.36%	-0.61%
SR3	1.67%	1.24%	1.61%	1.65%	0.76%	1.66%	0.74%	0.38%	1.26%
SR4	0.29%	0.52%	0.59%	0.99%	0.37%	0.29%	-0.22%	-0.18%	0.62%
SR5	0.21%	0.17%	0.23%	0.07%	0.22%	0.09%	0.36%	-0.04%	0.41%
SR6	0.37%	-0.50%	-0.54%	-0.67%	-0.66%	-0.29%	-0.10%	-0.10%	0.98%

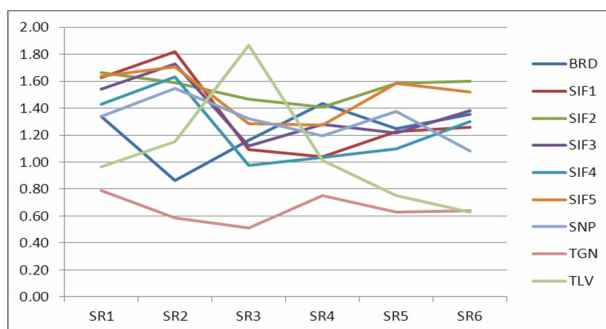
Table 4. Coefficient a_0 , return of market and variance of error on each sub-region (SR)

	α_{0j}	R_M	$\sigma^2(\varepsilon_i)$
SR1	-11.86%	0.89%	0.101923
SR2	6.94%	-0.44%	0.133241
SR3	-12.99%	0.91%	0.049174
SR4	-4.00%	0.03%	0.060762
SR5	-1.61%	0.48%	0.060749
SR6	2.74%	-0.01%	0.063653

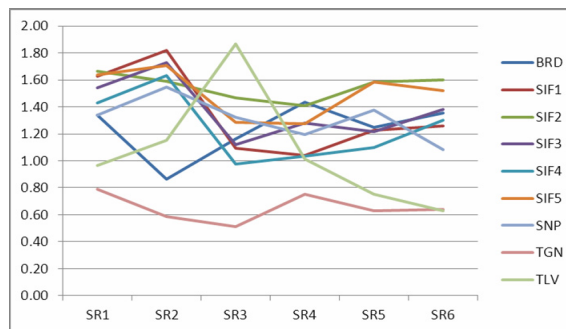
	α_{0j}	R_M	$\sigma^2(\varepsilon_i)$
SR1	-11.86%	0.89%	0.101923
SR2	6.94%	-0.44%	0.133241
SR3	-12.99%	0.91%	0.049174
SR4	-4.00%	0.03%	0.060762
SR5	-1.61%	0.48%	0.060749
SR6	2.74%	-0.01%	0.063653

Regresia spline asigura o variație mai redusă a erorilor valorilor variabilei dependente calculate, vizavi de valorile variabilei dependente observate empiric. În același timp se observă evoluția în timp a influenței fiecărei acțiuni asupra pieței, a evoluției rentabilității medii, precum și perioadele unde variația erorilor este mai mare – variație mai mare a prețurilor pe piață.

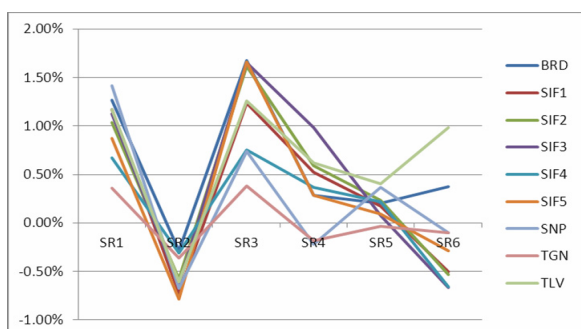
The spline regression ensure a smaller variance of error of values computed for the dependent variable, related to empirical observed values of dependent variable. Meanwhile, there can be observed the evolution over time of the influence of each share over market, of average return, and periods with larger variance of error – higher variation of prices on market.



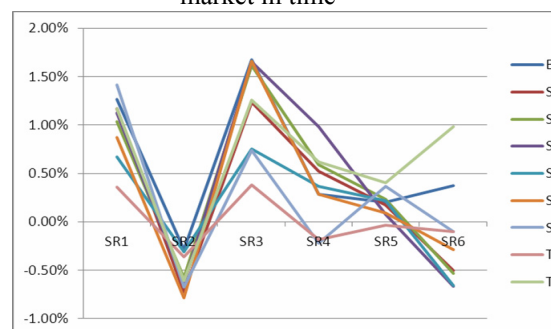
Graficul 1. Evoluția influenței acțiunilor asupra pieței în timp



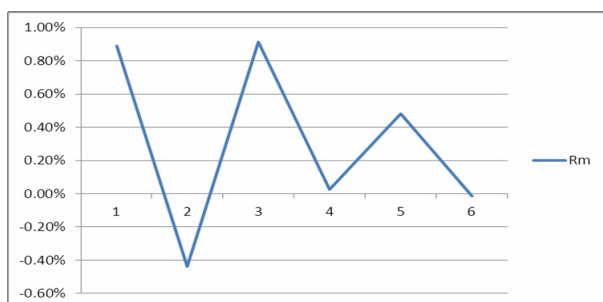
Graphic 1. Evolution of shares influence over market in time



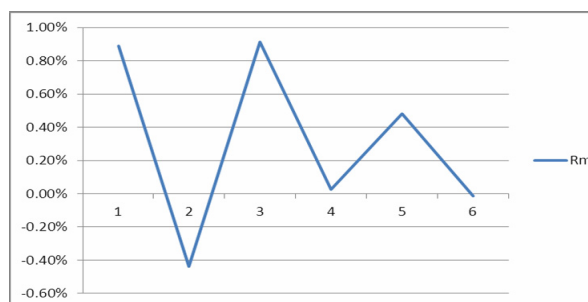
Graficul 2. Evoluția rentabilității medii a acțiunilor în timp



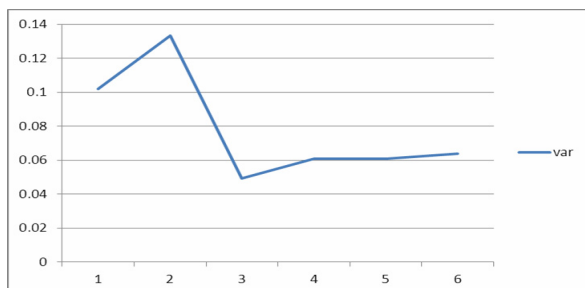
Graphic 2. Average return evolution for shares over time



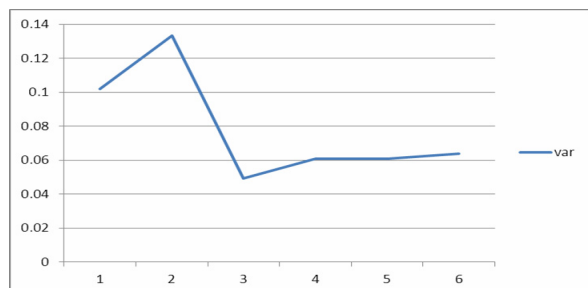
Graficul 3. Evoluția rentabilității medii a pieței pe subregiuni



Graphic 3. Average return evolution of market between sub-regions



Graficul 4. Evoluția variației erorii de previziune pe subregiuni



Graphic 4. Evolution of error variance between sub-regions

5. Concluzii

Regresiile spline dovedesc o putere mai mare de urmărire a trendului rezultat din datele empirice, evidențiind și evoluția în timp a fenomenelor, prezentând modul de evoluție a acțiunilor, a rentabilității pieței, a influenței acțiunilor asupra rentabilității pieței și a variației în timp a erorilor de aproximare.

Direcțiile viitoare de cercetare vor fi orientate spre crearea unor modele de regresii spline mai complexe cu variații ale erorilor mai reduse, determinarea subregiunilor relevante, precum și testarea modelelor create pe date extinse.

Bibliografie

Blundell Richard, Chen Xiaohong, Kristensen Dennis (2007), „Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves”, *Econometrica*, Vol. 75, No. 6, pg 1613-1669, 2007

Carson Richard T., Sun Yixiao, “The Tobit model with a non-zero threshold”, *Econometrics Journal*, vol. 10, pg. 488–502, 2007

Crainiceanu Ciprian M., Ruppert David, Wand M. P., „Bayesian Analysis for Penalized Spline Regression Using

5. Conclusions

Spline regressions prove to better follow the trends resulting from empirical data, highlighting also the evolution in time of phenomenon, presenting the way of evolution of shares, market returns, influence of shares evolution over market return and variance in time of errors.

Future research directions will focus on creating new spline regression models more complex, with error variance reduced, determining relevant sub-regions and testing created models on extended data.

Bibliography

Blundell Richard, Chen Xiaohong, Kristensen Dennis (2007), „Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves”, *Econometrica*, Vol. 75, No. 6, pg 1613-1669, 2007

Carson Richard T., Sun Yixiao, “The Tobit model with a non-zero threshold”, *Econometrics Journal*, vol. 10, pg. 488–502, 2007

Crainiceanu Ciprian M., Ruppert David, Wand M. P., „Bayesian Analysis for Penalized Spline Regression Using

WinBUGS”, *Journal of Statistical Software*, Volume 14, Issue 14, 2005

Engle Robert F., Rangel Jose Gonzalo, „The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes”, *The Review of Financial Studies*, vol. 21, no. 3, pg 1187 – 1222, 2008

Friedman, Jerome H., „Multivariate Adaptive Regression Splines”, *The Annals of Statistics*, vol 19, no,1, pg 1 – 67, 1991

Hallahan Charlie, „The Tobit model: an example of maximum likelihood estimation with SAS/IML”, *SAS Users Group International* 22, paper 293, pg. 1379 - 1384, 1997

Huang Jianhua Z., „Local asymptotics for polynomial spline regression”, *The Annals of Statistics*, 2003, vol. 31, no. 5, pg, 1600–1635, 2003

Hurley Deborah, Hussey James, McKeown Robert, Addy Cheryl, „An Evaluation of Splines in Linear Regression”, *SAS Users Group International* 31, California, Paper 147-31, 2006

Ranjan Priya, Tobias Justin L., „Bayesian inference for the gravity model”, *Journal of Applied Econometrics*, vol. 22, pg 817–838, 2007

Rodriguez German, „Research methods in Demography”, *Princeton University, Office of Population Research*, on-line: <http://data.princeton.edu/eco572/>, 2006

Schnedler, Wendelin, „Likelihood Estimation for Censored Random Vectors”, *Econometric Reviews*, vol. 24: 2, pg, 195 — 217, 2005

Thomas Lee, „On algorithms for ordinanz least squares regression spline fitting: a comparative study”, *Journal of Statistical Computation and Simulation*, vol. 72: 8, pg. 647-663, 2002

Tobin James, „Estimation of Relationships for Limited Dependent Variables”, *Econometrica*, vol. 26, no. 1, pp. 24-36, 1958

WinBUGS”, *Journal of Statistical Software*, Volume 14, Issue 14, 2005

Engle Robert F., Rangel Jose Gonzalo, „The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes”, *The Review of Financial Studies*, vol. 21, no. 3, pg 1187 – 1222, 2008

Friedman, Jerome H., „Multivariate Adaptive Regression Splines”, *The Annals of Statistics*, vol 19, no,1, pg 1 – 67, 1991

Hallahan Charlie, „The Tobit model: an example of maximum likelihood estimation with SAS/IML”, *SAS Users Group International* 22, paper 293, pg. 1379 - 1384, 1997

Huang Jianhua Z., „Local asymptotics for polynomial spline regression”, *The Annals of Statistics*, 2003, vol. 31, no. 5, pg, 1600–1635, 2003

Hurley Deborah, Hussey James, McKeown Robert, Addy Cheryl, „An Evaluation of Splines in Linear Regression”, *SAS Users Group International* 31, California, Paper 147-31, 2006

Ranjan Priya, Tobias Justin L., „Bayesian inference for the gravity model”, *Journal of Applied Econometrics*, vol. 22, pg 817–838, 2007

Rodriguez German, „Research methods in Demography”, *Princeton University, Office of Population Research*, on-line: <http://data.princeton.edu/eco572/>, 2006

Schnedler, Wendelin, „Likelihood Estimation for Censored Random Vectors”, *Econometric Reviews*, vol. 24: 2, pg, 195 — 217, 2005

Thomas Lee, „On algorithms for ordinanz least squares regression spline fitting: a comparative study”, *Journal of Statistical Computation and Simulation*, vol. 72: 8, pg. 647-663, 2002

Tobin James, „Estimation of Relationships for Limited Dependent Variables”, *Econometrica*, vol. 26, no. 1, pp. 24-36, 1958